

## Lexical Categories

We have three tests to determine whether a word is a noun, verb, or an adjective:

	meaning	morphology	syntactic distribution
noun (N)	objects, abstract concepts, individuals	plural, possessive	She can __ (it).
verb (V)	actions, sensations, states of being	past, progressive, 3SG present	The __ is/are here.
adjective (A)	descriptions of nouns	comparative, superlative	He/It seems rather __.

Looking at meaning is a quick way to estimate the lexical category of a word, but it's not always sufficient. For example, *see*, *perceive*, *behold*, *look*, *view*, and *glance* have very similar meanings, but *see*, *perceive*, and *behold* can only be used as verbs, while the other three can be used as verbs and as nouns.

This is where morphology comes in handy, but morphology can sometimes be irregular. For example, some nouns have irregular plurals (*men*, *oxen*) and some don't seem to even have plurals at all (*\*knowledges*, *\*furnitures*). Similarly, some verbs have irregular past tenses (*had*, *taught*), some of which are not distinct from their bare form (*put*, *set*).

Syntactic distribution can help confirm the lexical category of a word, but this has to be done carefully. Some of the resulting sentences used to test the lexical category may have odd meanings that make it difficult to tell if the problem is in the syntax or the semantics, so test sentences need to be constructed carefully.

In most cases, a combination of all three tests will usually successfully reveal the lexical category of a word, but of course, if a word can't be classified by the tests above, we need some other way of knowing what lexical category it is. For example, there aren't tests exactly like the ones above for prepositions.

## Phrase Structure

Once we know what lexical categories the words of a sentence belong to, we can compile the possible ways these lexical categories can be combined with each other to form grammatical sentences of the language. In class, we discovered that English sentences could have a variety of patterns, including:

N V	Boars ate.
Det N V	The boars ate.
A N V	Angry boars ate.
Det A N V	The angry boars ate.
N V N	Boars ate humans.
N V Det N	Boars ate the humans.
N V A N	Boars ate dead humans.
N V Det A N	Boars ate the dead humans.
etc.	

More generally, we can write  $S \rightarrow (\text{Det}) (\text{A}) \text{N V} ((\text{Det}) (\text{A}) \text{N})$ , meaning that a sentence can be composed of an optional determiner (words like *the*, *a*, and *some*), an optional adjective, a noun, a verb, and an optional sequence of an optional determiner, optional adjective, and noun.

There is a repeated pattern of (Det) (A) N occurring in two separate places in the sentence, with the three elements in exactly the same order, with exactly the same pattern of optionality. This is unlikely to be a coincidence, so if we want to have a rule that encodes this pattern. These three elements together constitute what is known as a noun phrase, and extracting this predictable NP pattern out into its own rule, we get:

$$\begin{aligned} S &\rightarrow NP V (NP) \\ NP &\rightarrow (Det) (A) N \end{aligned}$$

Further exploration of English sentences reveals repeated patterns for adjective phrases and for prepositional phrases, too, giving us a fuller set of phrase structure rules for English:

$$\begin{aligned} S &\rightarrow NP (PP) V (NP) (AP) (PP) \\ NP &\rightarrow (Det) (AP) N \\ AP &\rightarrow (Deg) A \\ PP &\rightarrow P NP \end{aligned}$$

## Constituents

So far, we have only come up with phrase structure rules to describe repeated patterns of words. It would be nice to find some independent support for these rules, some other evidence that shows that a noun phrase really is a separate unit, a **constituent**. There are at least three tests that show how constituents behave as a unit, while other strings of consecutive words do not. Not only do these tests confirm the existence of NPs, APs, and PPs, but also suggest the existence of VPs.

**Coordination Test:** Constituents of the same type can be joined together with conjunctions like *and* and *or*, while non-constituents cannot be joined in such a way.

Clark lied to [a bald billionaire] and [a blonde reporter].	NPs
Clark [lied to a bald billionaire] and [revealed his secret to a blonde reporter].	VPs
*Clark lied to [a bald billionaire] and [protect his secret].	NP ≠ VP

**Substitution Test:** Instead of repeating a complete constituent in a separate location, it may be replaced in its entirety by an appropriate substitute word (e.g., pronouns for NPs, *do so* for VPs, and *there* for locative PPs), while non-constituents (or incomplete constituents) cannot be substituted.

<u>A painting of an explosion</u> covered the floor, and <u>it</u> spurred everyone to action.	NP
* <u>A painting of an explosion</u> covered the floor, and <u>it</u> of the cop was on a canvas.	not complete

The older son <u>saved the world</u> if the younger son <u>did so</u> .	VP
*The older son <u>saved</u> the world if the younger son <u>did so</u> the cheerleader.	not complete

**Movement Test:** Under certain circumstances, a complete constituent can be moved to other parts of the sentence, sometimes with slightly different syntax, while non-constituents cannot be moved. There are a variety of movement tests (fronting and passives, for examples). Demonstrated here is pseudo-clefting, where the constituent to be tested appears at the beginning of the sentence, followed by a statement of the form *is/are who/what/where*, and the rest of the sentence, with the constituent no longer in its usual position. (Note that pseudo-clefting of VPs requires *do*-insertion.)

<u>The phone</u> is what Pam and Ryan answer __.	NP
* <u>Phone</u> is what Pam and Ryan answer the __	not complete

<u>Answer the phone</u> is what Pam and Ryan do __.	VP
* <u>Answer</u> is what Pam and Ryan do __ the phone.	not complete

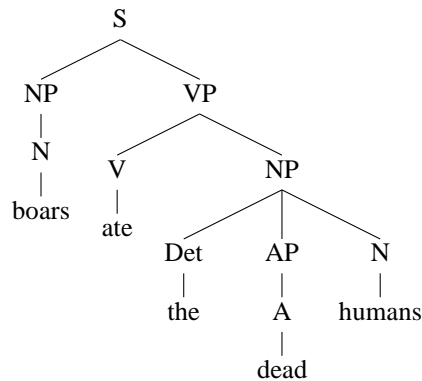
<u>A book of tasteless jokes</u> is what Michael has __.	NP
* <u>A book</u> is what Michael has __ of jokes.	not complete

## Phrase Structure Trees

With the discovery that VPs are constituents, and that NPs can contain PPs, we can revise our phrase structure rules to be:

$S \rightarrow NP VP$   
 $NP \rightarrow (Det) (AP) N$   
 $AP \rightarrow (Deg) A$   
 $PP \rightarrow P NP$   
 $VP \rightarrow V (NP) (AP) (PP)$

We can visualize the phrase structure of a sentence in a tree format, which is often easier to interpret than a purely linear bracketed, especially as the sentences get more complex.

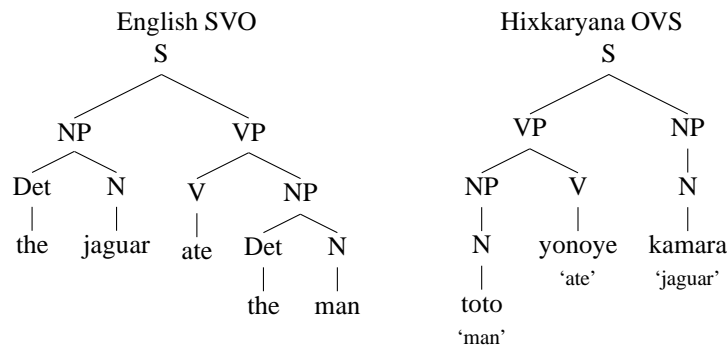


[S [NP boars<sub>N</sub>] [VP ate<sub>V</sub> [NP the<sub>Det</sub> [AP dead<sub>A</sub>] humans<sub>N</sub>] ] ]

## Universal Patterns

We find that this basic phrase structure seems to be the same across languages, except that the *order* is different. For example, NPs always seem to contain Det, AP, N, and PP, but not necessarily in the same order as in English. The phrase structure tree diagram can then be thought of as a mobile, with all of the parts strictly connected at the nodes in the tree, but with the individual branches able to freely swing around each other, but staying confined to their constituents.

For example, while English has SOV word order, some languages like Hixkaryana have OVS word order, with the main subcomponents of S or VP completely reversed, though the components themselves appear in the same constituents (the subject NP and the VP are both in directly under the S node, and the verb and object NP are both directly under the VP node).



‘The jaguar ate the man.’

‘The jaguar ate the man.’  
 \*‘The man ate the jaguar.’